 DOI: https://doi.org/10.12349/tie.v3i2.9143

Fault Traceability and Self-healing Strategies for Intelligent Control Systems Based on Causal Inference

Wenfu Zhang

Shandong College of Information Technology, Weifang, Shandong, 261061, China

Abstract

To address the core challenges of difficult fault tracing and imprecise self-healing in intelligent control systems, this paper proposes an integrated autonomous health management framework that combines causal inference with bio-immune inspiration. At the tracing level, an interpretable model based on temporal causal discovery (ExCTM) is constructed to achieve precise localization of root causes. At the self-healing level, a causal-knowledge-guided, bio-immune-inspired reinforcement learning method (BIH-RL) is designed to generate efficient and specific recovery strategies. Experiments on the Tennessee-Eastman process and a robotic arm system demonstrate that this framework significantly improves root cause localization accuracy and self-healing efficiency, providing a new pathway for building next-generation intelligent systems with "cognitive self-healing" capabilities.

Keywords

Causal Inference; Fault Root Cause Tracing; Self-Healing Control; Reinforcement Learning; Intelligent Systems

基于因果推理的智能控制系统故障溯源与自愈策略

张文甫

山东信息职业技术学院，中国 · 山东 潍坊 261061

摘　要

智能控制系统故障难溯源、自愈欠精准，本文融合因果推理与生物免疫启发提出一种一体化自主健康管理框架。从溯源层面上构建了基于时序因果发现的可解释模型（ExCTM），能够精准定位故障根本原因。在自愈层面上设计了因果知识引导的生物免疫启发式强化学习方法（BIH-RL），建立了一套高效、特异性的恢复策略。在田纳西-伊士曼过程和机械臂系统上的实验表明，能能显著提升根因定位准确率和自愈效率，为构建具备认知自愈能力的下一代智能系统提供了新途径。

关键词

因果推理；故障溯源；自愈控制；强化学习；智能系统

1 绪论

1.1 研究背景与意义

高端装备与复杂工业过程的智能控制系统需要极高的可靠性，在应对多变量耦合、故障传播隐蔽的系统时，传统故障诊断与容错方法可解释性差（黑箱模型）、关联非因果（难以定位根本原因）、诊断与恢复脱节。因此，需要有能够精准追溯故障根源并实现自主恢复的智能方法。

通过建立系统的结构因果模型（SCM），不仅能识别异常，更能推理故障的为什么与怎么办。本研究的目的是构建一个故障溯源与自愈一体化框架，同时在框架中融合因果推理与智能优化，让智能系统从感知诊断向认知自愈的自主健康管理模式转变。

【作者简介】张文甫（1988–），中国山东潍坊人，硕士，从事智能控制研究。

1.2 国内外研究现状综述

故障诊断：数据驱动方法（如深度学习）在异常检测方面表现出色，但缺乏因果解释性，易受伪相关干扰。现有因果发现方法（如 PC 算法）对动态时序系统的适应性不足。

自愈控制：主要以基于强化学习（RL）的方法为主，但其通常将自愈视为黑箱优化问题，搜索效率低且策略不可靠。

交叉研究：因果推理用于提升控制策略的泛化性，但系统性的从故障感知到自主恢复的全链条研究尚属空白。

核心问题：当前研究在故障精准溯源、诊断与自愈的闭环衔接以及自愈策略的可解释与自适应方面存在不足。

2 基于因果推理的智能控制系统故障溯源理论框架

2.1 智能控制系统故障的因果特性分析

我们将智能控制系统建模为一个结构因果模型（SCM）：M={U,V,F,P(U)}。

V 为观测变量（如传感器数据）

U 为外生隐变量

F 为定义变量间机制的结构方程集合。

系统的正常运行由该因果结构决定；故障本质是对此因果结构的干预，可分为两类：

对结构方程 F 的篡改（如执行器性能退化），导致变量间函数关系系统性改变，引发传导性故障。

对变量 V 的直接干扰（如传感器偏置），仅局部破坏数据，但错误会沿因果链传播，引发局部性故障。

表 2.1 故障的因果分类与影响

故障类型	因果干预对象	影响范围	示例
网络型故障	结构方程 F	传导性、全局性	催化剂失活、部件磨损
测量型故障	观测变量 V	局部性、传播性	传感器漂移、信号中断

2.2 面向故障溯源的因果推理基础

精准溯源依赖于两个核心因果概念：

干预（do- 演算）：P(Y ｜ do(X=x)) 表示将变量 X 强制设为 x 后 Y 的分布。用于预测主动措施的效果。

反事实推理：回答“如果当时…，那么会…”的问题，是定位实际原因（Actual Cause）的关键。我们采用 Halpern-Pearl 框架的变体，认为变量 X=x 是异常 ϕ 的一个实际原因，需满足：

事实性：X=x 且 ϕ 发生。

必要性：存在 X 的正常取值 x′，使得若 X 取 x′（反事实干预），则 ϕ 不会发生。

最小性：不存在更小的变量子集满足上述条件。

通过计算变量的必要性概率（PN）与充分性概率（PS），可确定根本原因集（RCS）。

2.3 融合深度学习的可解释因果溯源模型设

本文提出可解释因果溯源模型（ExCTM），其核心是学习潜在因果变量 $Z_{1:T}$ 的动态，并监测其因果机制的异常。模型架构如图 1 所示，包含两大模块：

2.3.1 时序表征与潜在因果发现（TCN-VAE-MHA 模块）

该模块从原始高维时序数据 $X_{1:T}$ 中提取低维潜在因果变量 $Z_{1:T}$。

TCN 层：利用膨胀因果卷积捕获长期依赖。

VAE 层：推断后验分布 $q_\phi(Z \mid X)$，迫使 Z 学习解耦的、稳定的数据生成因子。

MHA 层：聚焦关键异常时刻与变量。

因果图学习：Z 应用连续优化方法，学习描述其相互作用的有向无环图（DAG）G_Z。

2.3.2 动态异常检测与根因定位（ICODE 模块）

把潜在变量的动态用神经常微分方程（Neural ODE）参数化，函数结构受 G_Z 约束：

$$\frac{d\mathbf{Z}(t)}{dt}=f_\theta(\mathbf{Z}(t),t;G_Z)$$

通过计算因果残差 $\varepsilon t=Z_t-\hat{Z}_t$ 和参数偏移量 $\Delta\theta_t$ 来检测故障。

在线检测到故障则启动根因定位：

1. 识别首批异常变量集合 A。

2. 在因果图 G_Z 上从 A 反向追溯父节点，并为每条边 $(Z_j \to Z_i)$ 计算反事实贡献度分数（CCS）：

$$CCS_{j\to i}=\left| E\left[Z_i \mid do(Z_j=z_j^{obs})\right]-E\left[Z_i \mid do(Z_j=z_j^{normal})\right] \right|$$

3. 递归选择 CCS 最大的父节点，结合轻量化反事实推理评估其必要性与充分性，输出最终的根本原因集（RCS）。

ExCTM 最终将定位的潜在根因映射回原始观测空间，并且输出可视化的因果子图，提供直观解释。

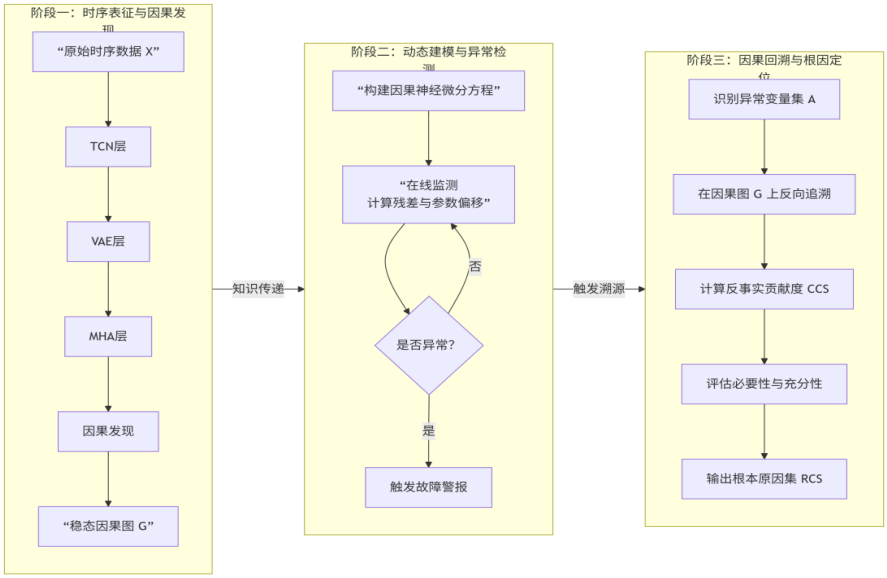

图 1ExCTM 模型架构图

3 基于因果知识引导的智能自愈策略生成与优化

3.1 自愈控制的基本原理与生物免疫启示

系统在故障后能自主调整结构或参数，最终恢复功能是自愈控制的核心。实现高智能的自愈需系统具备感知、诊断、规划与学习能力。可以借鉴自然界中的生物免疫系统运作方式。

免疫系统具备特异性、记忆性与自适应性，是一个多层防御网络。其与智能自愈控制系统的功能映射关系，为核心框架设计提供了根本性启示（表 3.1）。

表 3.1 生物免疫系统与智能自愈控制系统功能映射

生物免疫系统	智能自愈系统	核心启示
抗原识别与呈递	故障检测与因果溯源	精准识别“非己”（故障）并定位关键特征（根因）。
适应性免疫应答	智能自愈策略生成	针对特定“抗原”产生高效“抗体”（修复策略）。
免疫记忆	策略优化器与经验池	记住有效策略，加速未来应对。
免疫调节	稳定性与代价约束	防止过度反应，维持系统稳态。

3.2 因果知识引导的自愈动作空间构建

传统的无模型强化学习（RL）在庞大的全局动作空间中探索，效率低下。第二章输出的因果溯源结果，为动作空间提供了精准的手术刀，实现动态、情境式的空间裁剪。

设全局动作为 A_{global}，根因集为 V_{rc}，因果图为 G，故障类型为 FaultType。我们定义一个映射函数 $\Gamma(\cdot)$，输出候选动作空间：

$$A_{candidate}=\Gamma(A_{global},\mathbf{V}_{rc},G,\text{FaultType})$$

该函数遵循以下原则：

聚焦根因：若根因为传感器故障，则 $A_{candidate}$ 聚焦于信号重构、冗余切换等操作，排除无关的物理调节动作。

利用因果图：允许在根因的下游关键节点进行补偿性干预，以阻断故障传播，拓展修复维度。

预判动作类型：对“网络型”故障，优先参数调整或部件切换；对“测量型”故障，优先信号校正。此举可大幅减少无效探索。

3.3 基于生物免疫启发式强化学习（BIH–RL）的自愈策略优化

基于以上原理，本文提出生物免疫启发式强化学习（BIH-RL）框架，其架构如图 3.1 所示，采用分层决策模式。

3.3.1 状态表征与免疫感知层

系统状态增强为 $s_t^+=[s_t,I_{rc},G_{sub},L_{fault},\epsilon_t]$ 其中融合了原始观测 s_t 及因果溯源提供的根因标识 I_{rc}、因果子图 G_{sub}、故障类型标签 L_{fault} 和因果残差 ϵ_t，赋予智能体深层的故障语义感知。

3.3.2 分层策略网络

高层策略（免疫中枢）π^{high}：接收 s_t^+，输出抽象的自愈目标模式 g_t（如“参数微调”、“控制律切换”），制定宏观修复路径。

底层策略（效应器）π^{low}：接收 s_t^+ 和 g_t，在 candidate 空间内输出具体动作 a_t，专业化地执行高层目标。

3.3.3 因果融合的奖励函数设计

奖励函数是引导智能体学习“治本”策略的关键，由三部分组成：

$$r_t=w_1R_{\text{perf}}+w_2R_{\text{causal}}+w_3R_{\text{stab}}$$

R_{perf}：基于系统主要性能（如跟踪误差）恢复的传统奖励。

R_{causal}：核心创新奖励。

R_{stab}：惩罚过大或频繁的动作，确保修复过程平滑稳定。

3.3.4 免疫记忆与高效探索

设计分类免疫经验回放池，依据“故障类型 - 根因”对经验分类存储。遇到新故障时，优先从同类经验中采样，实现“记忆”快速唤醒。探索过程被限制在 $A_{candidate}$ 内，并随学习进程聚焦于高潜力方向，实现引导式高效探索。

4 实验设计

本章在田纳西 - 伊士曼（TE）化工过程与双连杆机械臂仿真平台上验证框架有效性。核心是对比所提框架（ExCTM+BIH-RL）与以下基准方法：

传统方法：主成分分析（PCA）贡献图法。

先进数据驱动方法：长短期记忆自编码器（LSTM-AE，用于诊断）、深度确定性策略梯度（DDPG，用于自愈）。

4.1 消融实验方法

ExCTM w/o Causal：移除因果约束的溯源模型。

BIH-RL w/o Immune-Memory：移除免疫记忆的自愈策略。

评价指标聚焦于诊断溯源性能（根因定位准确率 RCA、诊断延迟 DD）与故障自愈性能（修复成功率 RSR、平均修复时间 MRT、干预成本 IC）。

4.2 案例一：TE 化工过程多重故障并发溯源与自愈

场景设置：在 TE 过程中同时引入故障 4（反应器冷却水入口温度阶跃上升）与故障 9（D 进料温度随机波动），构成并发干扰场景。

结果与分析：关键结果对比如表 4.1 所示。本文框架的 ExCTM 模型展现出最高的根因定位准确率（92.5%），能有效区分并定位首要根因（冷却水故障），而非像 PCA 和 LSTM-AE 那样将主要贡献指向中间变量。

在自愈方面，BIH-RL 策略凭借因果引导，将动作精准聚焦于调整冷却水系统，实现了 100% 的修复成功率、最短的平均修复时间（65 分钟）和最低的干预成本。消融实验证明，因果约束与免疫记忆机制对提升性能均有显著贡献。

表 4.1 TE 过程并发故障诊断与自愈结果综合对比

方法	根因定位准确率 (RCA%)	诊断延迟 (DD/min)	修复成功率 (RSR%)	平均修复时间 (MRT/min)	干预成本 (IC)
本文框架 (ExCTM+BIH-RL)	92.5	28	100	65	1.0(基准)
PCA 贡献图法	65.0	45	不涉及	不涉及	不涉及
LSTM-AE 诊断模型	78.3	35	不涉及	不涉及	不涉及
DDPG 自愈策略 (无因果引导)	不涉及	不涉及	70	>120	1.9
ExCTM w/o Causal (消融)	71.6	32	不涉及	不涉及	不涉及
BIH-RL w/o Immune-Memory (消融)	92.5	28	95	78	1.1

4.3 案例二：机械臂关节突发故障在线自愈

场景设置：在机械臂轨迹跟踪任务中，于 t=5s 时模拟关节 2 电机扭矩突发下降 40% 的执行器故障。

结果与分析：自愈性能对比如表 4.2 所示。本文框架通过 ExCTM 准确定位“扭矩增益下降”这一根因，并由 BIH-RL 生成针对性的扭矩重分配与参数补偿策略，从而实现了最小的跟踪误差和最高的修复成功率。相比之下，无模型引导的 PPO 算法表现出振荡和不稳定的恢复过程。

表 4.2 机械臂关节故障自愈性能对比

方法	最大跟踪误差 (恢复期)/rad	恢复后稳态误差 /rad	修复成功率 (RSR%)
本文框架	0.15	0.008	100
模型参考自适应控制 (MRAC)	0.25	0.025	100
近端策略优化 (PPO)	0.35	0.050	80

4.4 综合讨论

实时性：在实验硬件上，ExCTM 推理与 BIH-RL 决策的总时间远小于系统采样间隔，满足实时性要求。

局限性：对引入全新未知因果关系的故障泛化能力可能有限、对极端快变故障的响应存在固有延迟、框架性能在一定程度上依赖于初始因果知识的准确性。

通过两个典型案例的系统验证，本章证实了所提框架在提升故障溯源精准度、自愈策略效率与经济性方面的显著优势，为实现智能系统“认知自愈”提供了有效的技术途径。

5 总结与展望

5.1 研究工作总结

本研究针对智能控制系统故障精准溯源与自主恢复的挑战，提出了一个融合因果推理与生物免疫启发的一体化自主健康管理框架。主要工作与结论如下：

理论框架层面：基于结构因果模型（SCM）形式化定义了系统故障，构建了可解释时序因果溯源模型（ExCTM），实现了对动态系统根本原因变量的精准定位。

方法创新层面：提出了生物免疫启发式强化学习（BIH-RL）自愈策略生成方法，利用因果知识引导策略搜索与优化，实现了高效、特异性的故障恢复。

实验验证层面：在 TE 化工过程与机械臂平台上的实验表明，本框架能有效提升根因定位准确率与自愈效率，验证了其先进性与实用性。

5.2 未来研究展望

本工作为智能系统自主管理提供了新思路，未来可从以下方向深化：

理论层面：探索小样本下的稳健因果发现方法，以及处理未观测混杂和循环因果的更通用推理理论。

技术层面：研究面向系统级故障的多智能体协同自愈，并结合生成式 AI 与世界模型增强策略的泛化与预训练能力。

应用层面：推动因果数字孪生验证平台的标准化，并研究算法的轻量化部署与人机协同交互机制，以促进技术落地。

参考文献

[1] PEARL J.Causality:Models,Reasoning,and Inference[M].2nd ed. Cambridge:Cambridge University Press, 2009.

[2] ZHENG X,DAN C,ARAGAM B,et al. Learning Sparse Nonparametric DAGs [C]// Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics. 2020:3414-3425.

[3] 张化光，王迎春.智能故障诊断与自愈控制 [M].北京：科学出版社，2018.

[4] LEVINE S, FINN C, DARRELL T, et al. End-to-End Training of Deep Visuomotor Policies [J]. The Journal of Machine Learning Research, 2016, 17(1): 1334-1373.

[5] HAO J,CHEN H, ZHANG R, et al. Biologically Inspired Self-Healing Control Systems:A Review [J]. IEEE Transactions on Neural Networks and Learning Systems,2022,33(12): 6987-7001.

[6] 刘强，周东华.数据驱动的工业过程故障诊断研究综述 [J]. 自动化学报，2016,42(9):1285-1299.